

## 逆翻訳による言い換え生成を用いた協調学習自動評定の性能改善

福田 治輝<sup>†</sup> 綱川 隆司<sup>†</sup> 大島 純<sup>†</sup> 大島 律子<sup>†</sup> 西田 昌史<sup>†</sup> 西村 雅史<sup>†</sup>  
静岡大学 情報学部<sup>†</sup>

## 1. はじめに

協調学習では、複数の学習者がグループ活動を通して協力し、意見を交わしながら課題に向き合う。協調学習内で行われる調整活動プロセスにおける調整能力の評価手法として CSSER (Collaboration Scenario-based Scale for Emotion Regulation) と呼ばれる手法が提案されている[1]。CSSERではあるシナリオにおいて登場する学習者の思考や発言を表す状況に応じた内容のテキストを記入させ、評価を行っている。グループ編成に CSSER の結果を活用するには短期間で評定作業を済ませる必要があるが、ループリックに基づいて人手で評定しているため、学習者数の多い場合は実用的でない。

本研究では、協調学習における調整能力の評価手法である CSSER において、即時的なフィードバックを返すことを可能にするための評定作業の自動化手法を改善するため、逆翻訳を用いた言い換えテキストをトレーニングデータに追加する方法を提案しその効果の評価した。

## 2. CSSER

Oshima ら[1]により作成された4種類の CSSER シナリオでは被験者自らが参加する協調学習場面が再現され、同じグループに属する登場人物同士の会話を描いたコマが続く形でストーリーが展開されている。シナリオの最後には、被験者自身を表現したキャラクターの上に空欄の吹き出しが設けられたコマが用意されており、その状況で考えている内容（以下、「心の声」とする）とその状況で発言する内容（以下、「発言」とする）を想定し、記述させる。それぞれのシナリオで記述された「心の声」と「発言」に対し、社会認識的側面と社会感情的側面の2側面から既定のループリックに従って1~5点の評定を与える。つまり、学習者の CSSER 評定は、4シナリオ\*2つの記述項目（心の声、発言）\*2つの評価側面（社会感情的側面、社会認識的側面）の計16の評定となる。

Improvement of automatic text scoring for collaborative learning using back-translated paraphrases

<sup>†</sup> Haruki Fukuda, Takashi Tsunakawa, Jun Oshima, Ritsuko Oshima, Masafumi Nishida, Masafumi Nishimura  
Faculty of Informatics, Shizuoka University

## 3. 提案手法

## 3.1 BERT

BERT は自己注意機構のみを使用している Transformer[4]をベースとした汎用言語表現モデルである。大規模なテキストコーパスで事前学習を行い、各タスクに対してファインチューニングすることにより、多くの自然言語処理タスクにおいて有効に動作する。

## 3.2 Google 翻訳による逆翻訳を用いた言い換えテキストの生成

Google 翻訳 API の v3 NMT モデルを利用し、CSSER の各回答テキストを日本語から韓国語、中国語（簡体字）、英語へ翻訳を行い、再度それぞれ日本語へ翻訳を行うことで元の回答テキストの言い換えテキストを生成する。

ここで、日本語として自然な逆翻訳テキストを選別するため、BERT as Language Model を利用し逆翻訳された個々のテキストに対して Perplexity を算出し、Perplexity が相対的に小さいテキストを言い換えテキストとして用いる。

## 3.3 適用方法

日本語 Wikipedia 記事（約 1800 万文）によってプレトレーニングされたモデル[5]を用いて、BERT への入力を CSSER の「心の声」または「発言」のテキスト及び生成した言い換えテキスト、出力を社会認識的側面または社会感情的側面の1~5点の評定とし、それぞれクラス分類を行うようにファインチューニングを行う。1 から 5 の各評定について得られた尤度が最大である評定を、自動評定の結果として採用する。

## 4. 評価

## 4.1 実験設定

グループワークが主となる静岡大学の演習系講義の 2015 年度から 2017 年度を受講生に対して、初回講義の冒頭および最終回講義の冒頭の 2 回 CSSER を実施した。「心の声」と「発言」合わせて 10293 件の回答が得られた。これらのシナリオ別の回答データおよび人手での評定を実験対象データとする。

実験対象となるデータを 3:1:1 の割合でランダムにトレーニングセット、バリデーションセット、

テストセットに分割する。トレーニングセットから生成した言い換えテキストのうち Perplexity の値の小さいものから、25%、50%、75%、100%の割合でトレーニングデータに含める場合と、全く含めない場合で評価実験を行った。

評価指標として、Cohen の一致係数（以下  $\kappa$  係数）及び F 値のマイクロ平均を用いる。BERT の主なパラメーターとして、語彙サイズ（サブワード含む）32,000、バッチサイズ 32、エポック数 3 を用いた。

#### 4.2 結果

シナリオ別に行った実験結果の平均を表 1 に示す。社会感情的側面，社会認識的側面のいずれの側面においても言い換えテキストをトレーニングデータとして含めた場合の方がわずかに高い推定精度となった。

発言の社会認識的側面以外の結果から、言い換えテキストを全てトレーニングデータに含める場合よりも 50%または 75%のトレーニングデータを含めた場合の方が高い推定精度となっていることがわかる。100%トレーニングデータに追加した場合よりトレーニングデータが少ないにもかかわらず精度が上がっていることから、Perplexity の大きいテキストを取り除くことは有効であると考えられる。

#### 5. おわりに

CSSER では人手で評定することにより評価を行っているが、人手による作業は精度が高いものの時間がかかりすぎ、即時的なフィードバックを返せないという大きな欠点がある。この評定作業を自動化する手法の改善方法として、逆翻訳を用いた言い換えテキストをトレーニングデータに追加することによる評定精度の改善方法を検討し、評価した。Perplexity によってフィルタリングを行った言い換えテキストをトレーニングデータに導入することで、わずかに効果が認められた。今後はモデルの改善等により評定精度のさらなる向上を目指し、人手での評定と併用する場合、類似テキストの提示等の評定支援の機能に関しても今後検討していく。

表 1 トレーニングデータに追加した言い換えテキストの割合（Perplexity の昇順）ごとの評定性能

		発言		心の声	
		$\kappa$ 係数	F 値	$\kappa$ 係数	F 値
社会感情的側面	0%	0.348	<b>0.570</b>	0.242	0.516
	25%	0.344	0.562	0.257	<b>0.522</b>
	50%	<b>0.358</b>	<b>0.570</b>	0.256	<b>0.522</b>
	75%	0.348	0.567	<b>0.258</b>	0.515
	100%	0.349	0.567	0.257	0.511
社会認識的側面	0%	0.349	0.579	0.182	0.531
	25%	0.385	0.592	0.218	0.533
	50%	0.371	0.585	0.221	0.559
	75%	0.388	0.590	<b>0.244</b>	<b>0.560</b>
	100%	<b>0.395</b>	<b>0.597</b>	0.230	0.546

#### 謝辞

本研究は JSPS 科研費 JP19H01714 の助成を受けたものである。

#### 参考文献

- [1] R. Oshima et al., “Collaboration Scenario-based Scale for Emotion Regulation: Measuring Learners’ Agency to Regulate Own, Others’ and Group Emotions,” in Proc. of EdMedia + Innovate Learning 2015, pp. 796–801, 2015.
- [2] 神戸優 et al., “学習調整能力評価サイトの Web デザイン検討,” 日本教育工学会, pp. 391–392, 2018.
- [3] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv:1810.04805 2018.
- [4] A. Vaswani et al., “Attention Is All You Need,” in Proc. of the 31st International Conference on Neural Information Processing Systems, pp. 6000–6010, 2017.
- [5] 柴田知秀 et al., “BERT による日本語構文解析の精度向上,” 言語処理学会, pp. 205–208, 2019.